

The Interplay Between Theoretical and Experimental Probability: Beyond “Sample Size Matters”

Michael Meagher, Brooklyn College - CUNY

This article presents a series of class activities that develop an extended examination of the interplay between theoretical and experimental probability. In some cases an experiment can be used to confirm a theory and in other instances it can be used to develop a theory. Examples include coin-tossing, a dice game, and cup dropping with Monte Carlo approaches to probability discussed. This set of activities could be used with preservice teachers to improve their content knowledge in the area of probability as well as provide both a model of inquiry-based approaches and a forum for discussing pedagogical techniques involving hands-on activities. They could also be used in middle school classrooms to help students experience the power of probability experiments in examining real-life phenomena.

Introduction

NCTM Standards call for students in Grades 6 -8 to “formulate questions, design studies, and collect data about characteristics within one population” (NCTM, 2000). Modeling is an important aspect of this since, as Konold, Harradine, and Kazak (2007) argue, “Through modeling, students encounter and use probability in virtually the same way as practitioners do, with the purpose of better understanding some real phenomenon. Additionally, this approach frequently requires that students articulate their informal theories about probability and then put them to the test” (p. 217). In a study with teachers of statistics, Liu and Thompson (2007) found the probability and statistics knowledge of teachers highly compartmentalized and stressed the importance of teachers being engaged in actual or simulated experiments to develop their knowledge of probability. Finally, the new Common Core State Standard Initiative (2010) emphasises mathematical modeling, and the use of mathematics and statistics to analyse empirical situations.

A typical example of modeling seen in middle school classrooms across the country is for students to toss a fair coin 10 times and record the number of heads and the number of tails. Typically, the data is aggregated and students observe that the sample from the entire class is closer to 50 – 50 than most of the individual student samples. This activity can be the basis for many discussions such as:

- the likelihood of events
- the difference between one student’s results and the aggregation of the group
- whether it should ever be exactly 50 – 50
- the fact that the Law of Large Numbers is not a limiting theorem (i.e., you do not necessarily get closer to 50 – 50 by increasing the sample size in a limiting sense)
- the importance of sample size.

However, when this activity is conducted there is often no viewpoint beyond “sample size matters” and no trajectory embedded in the activity to use this one activity as a

The new Common Core State Standard Initiative (2010) emphasises mathematical modeling, and the use of mathematics and statistics to analyse empirical situations.

basis for doing further activities that could build deeper conceptual understanding. It is also often not the basis for a larger discussion about the interplay between theoretical probability and experimental probability. Coin tossing is an example where we know what's supposed to happen. What if we don't know what's going to happen or, furthermore, if we can't know what's going to happen? What is the role of probability in these cases, and what is the interplay between theoretical probability and experimental probability? These are important questions for teachers of mathematics to consider and consideration of such questions is at the heart of addressing the concerns of Konold, et al. (2007), and Liu and Thompson (2007) discussed above.

An interesting way in which to use the fair coin tossing activity is for it to be the first in a series of activities developing an extended examination of the interplay between theoretical and experimental probability. In some cases an experiment can be used to confirm a theory; in other instances it can be used to develop a theory. In the following paragraphs, I describe five activities that give preservice teachers the opportunity to explore the use of experiments to confirm or develop theories. Specifically, the five activities address the following five scenarios:

- (i) We have a theory and evidence from an experiment supports the theory
 - (ii) A more complicated example where we have a theory supported by evidence
 - (iii) We have a theory that is not supported by the evidence, but the theory can be revised
 - (iv) We do not have a theory, but we can build a convincing conjecture from the evidence
 - (v) We do not have a theory, and it is difficult to see how one might be built.
- Most, if not all, probability activities fall

into one of these categories, but it could be argued that the only one commonly addressed is the case of confirming a theory, and then only for the purpose of arguing that a large sample size will make experimental results close to theoretical results.

Each of the following activities took place with a group of twenty-one preservice teachers taking a Master's course in middle school mathematics teaching.

Part I: We have a theory that is supported by the evidence

In the first activity, the preservice teachers engaged in a standard activity whereby each of them tossed a coin 10 times and recorded the number of heads and tails. Before beginning the experiment, the expectations were discussed and the preservice teachers proposed that the number of heads should, in theory, be equal to the number of tails, and allowing for variation the experimental data should reflect this. The individual results varied from one instance of 2 – 8 to a couple of 5 – 5 splits, but the totals of 103 heads and 107 tails provided strong evidence that the theory of a 50 – 50 split is correct and that they used fair coins.

Part II: Another theory supported by the evidence

For the second activity, the preservice teachers were assigned to groups of three to play a game. Each group was given a six-sided die and a chip. Each group then constructed a playing board that consisted of a number line ranging from -2 to +2. The chip was placed at zero to begin the game. Two of three preservice teachers were players, one assigned High (4, 5, 6) and one assigned Low (1, 2, 3). The third person was the die thrower and chip mover. The die was thrown and, if a 4, 5, or 6 came up the counter was moved one unit in the

An interesting way in which to use the fair coin tossing activity is for it to be the first in a series of activities developing an extended examination of the interplay between theoretical and experimental probability.

positive direction. The chip was moved one unit in the negative direction if a 1, 2, or 3 was thrown. The game finished when the chip landed at one of the extremes (-2 or +2). A sample game is shown in Table 1.

Table 1 Sample game

Number Thrown	Beneficiary	Chip Position
3	L	-1
5	H	0
3	L	-1
6	H	0
6	H	+1
4	H	+2

The game would now be over and High would win. Before play started the preservice teachers argued that since the $P(1, 2, \text{ or } 3) = P(4, 5, \text{ or } 6) = 1/2$ each player is equally likely to win and that in 10 games the theoretical expectation is that each player would win 5 games. Each group of preservice teachers then played 10 games and recorded the number of wins for High and Low. The results can be seen in Table 2.

The preservice teachers were satisfied that the aggregate results: 38 (54%) wins for High and 32 (46%) wins for Low, supported the 50 – 50 theory and argued that a larger sample size would likely provide even stronger evidence.

Another possible direction to go in at this

Table 2 Data from sets of 10 die toss games with High (4, 5, or 6)

	Positive wins	Negative wins
Group 1	3	7
Group 2	6	4
Group 3	4	6
Group 4	7	3
Group 5	5	5
Group 6	7	3
Group 7	6	4
Total	38 (54%)	32 (46%)

point is to analyze game length noting that a given game could, in theory, last forever. However, we will continue on a trajectory of examining the interplay between theory and experiment.

Part III: We have a theory but it is not supported by the evidence

The third activity was a variation on the game played in Part II above. This time High was associated with a throw of 3, 4, 5, or 6, and Low with 1 or 2. The preservice teachers discussed their expectations of this game and, with very few dissenting voices, agreed that the number of wins would reflect the probabilities of throwing High or Low on a given roll and that High would win more than 2/3.

The preservice teachers then played the game with the results shown in Table 3. These results, 129 (76%) wins for High and 41 (24%) wins for Low, prompted a great deal of discussion among the preservice teachers with some claiming that the theory was incorrect and others claiming that the results were simply evidence of experimental variation. The preservice teachers were then challenged to do some calculations to develop a theory. There was some discussion in the groups about the fact that the game could go on

Table 3 Data from sets of 20 die toss games with High (3,4, 5 or 6)

	+ve wins	-ve wins
Group 1	18	2
Group 2	12	8
Group 3	17	3
Group 4	16	4
Group 5	16	4
Group 6	15	5
Group 7	15	5
Group 8	12	8
Group 9	8	2
Total	129 (76%)	41 (24%)

The preservice teachers discussed their expectations of this game and, with very few dissenting voices, agreed that the number of wins would reflect the probabilities of throwing High or Low on a given roll

for an infinite length of time (e.g. with rolls of 1, 6, 1, 6, 2, 6, . . . and were not confident about calculating an associated probability. However, many groups soon realized that games of finite length would have particular probabilities associated with them and proceeded to calculate those probabilities. The problem was analyzed in Table 4 where H stands for a throw of 3, 4, 5, or 6 and L stands for a throw of 1 or 2. The preservice teachers argued that a game could not last for three moves since a three move game would require the chip to be at -1 or +1 after two moves, each of which

would require the chip to be at 0 after one move, which is impossible. A conjecture emerged at this point that each way for H to win is mirrored by a way for L to win with the $P(H \text{ wins})$ larger than $P(L \text{ wins})$ by a factor of 4. The first steps towards a solution are shown in Figure 1.

Table 4 Games of length 2

	Game moves	Probability
H wins	HH	$\frac{4}{6} \times \frac{4}{6} = \frac{16}{36} = \frac{4}{9}$
L wins	LL	$\frac{2}{6} \times \frac{2}{6} = \frac{4}{36} = \frac{1}{9}$

Table 5 Games of length 4

	Game moves	Probability
H wins	HLHH	$\frac{4}{6} \times \frac{2}{6} \times \frac{4}{6} \times \frac{4}{6}$ $= \frac{128}{1296} = \frac{8}{81}$
	LHHH	$\frac{2}{6} \times \frac{4}{6} \times \frac{4}{6} \times \frac{4}{6}$ $= \frac{128}{1296} = \frac{8}{81}$
L wins	LHLL	$\frac{2}{6} \times \frac{4}{6} \times \frac{2}{6} \times \frac{2}{6}$ $= \frac{32}{1296} = \frac{2}{81}$

Many groups soon realised that games of finite length would have particular probabilities associated with them and proceeded to calculate those probabilities.

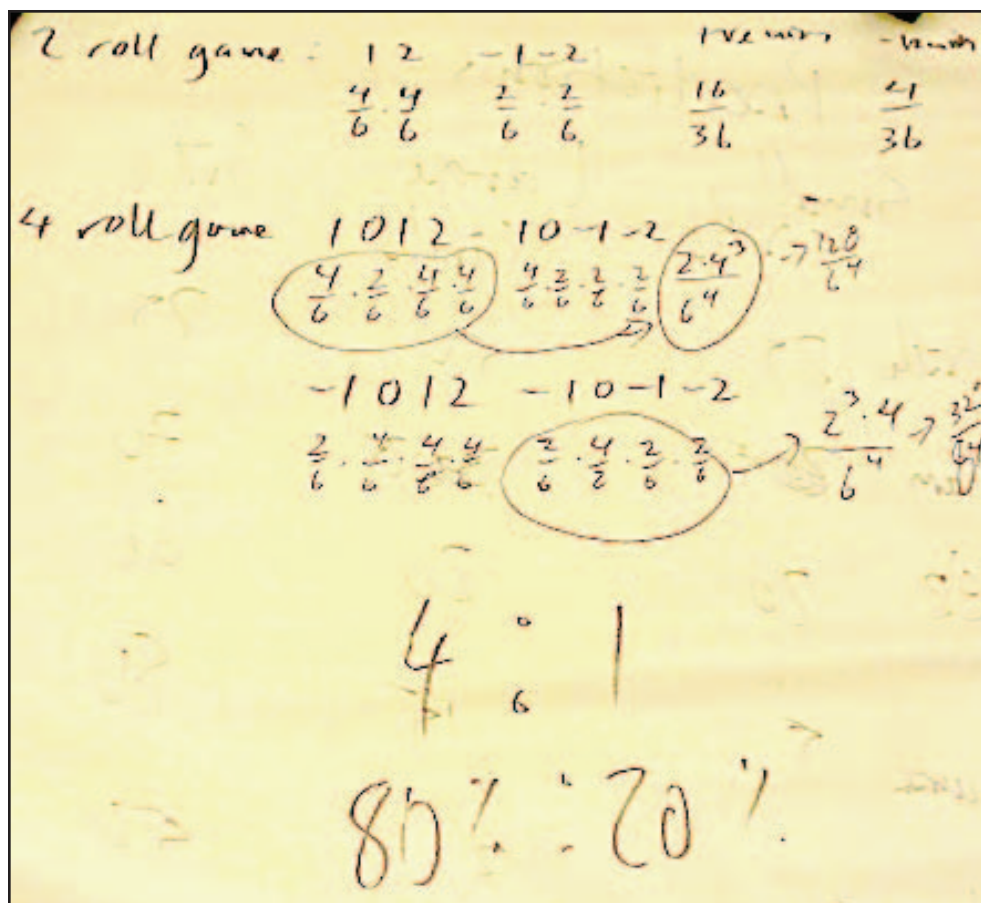


Fig 1 Calculations of theoretical probability for the “unfair” game

The preservice teachers went on to do the more calculations and, although they did not do a full analysis treating infinite cases they convinced themselves that the theory that High should win 80% of the time was correct. It was agreed that the data collected, 129 (76%) wins for High and 41 (24%) wins for Low, was strong evidence to support the theory. (Again, other avenues for exploration present themselves such as skewing the odds even more in favor of High (1, 2, 3, 4, 5) with Low (1) or changing limits of the board to +3 and -3.)

Part IV: We don't have a theory but can attempt to build one from the evidence

For the fourth activity, preservice teachers dropped plastic cups, held mouth upwards, from 1 foot, as measured by a ruler, and recorded the number of times that the cup settled upright mouth up, upright mouth down, and on its side.



Fig 2 Upright mouth up (small circle); upright mouth down (large circle); on side

The preservice teachers discussed possible results and even listed factors that would influence the probabilities, such as height of drop, diameter of the cup's mouth, and "bounce" of the cup. Predictions/guesses are shown in Table 6.

Table 6 Predictions for cup dropping (as %)

	Small Circle	Large Circle	Side
Nick	25	50	25
Mary	5	25	70
Elz	20	20	60
Lauren	5	15	80
Alex	20	30	50

The preservice teachers then worked in pairs to drop a cup 25 times and record the results which were: Small Circle 7 (3.5%), Large Circle 22 (11%), and Side 171 (85.5%) with the full results in Table 7.

Table 7 Data from sets of 25 cup drops

	Large O	Small O	Side
S & G	3	0	22
G & P	3	1	21
C & F	0	2	23
B & I	3	0	22
T & R	2	1	22
K & J	2	1	22
M & J	4	1	20
D & N	5	1	19
Total	22 (11%)	7 (3.5%)	171 (85.5%)

In the discussion that followed this experiment, the preservice teachers were surprised at the stability and consistency of the results across groups. This led to a discussion about the importance of dropping the cup from 1 foot and dropping the cup mouth up. The experiment was repeated first from the same height with the mouth of the cup down, and then with the cup dropped with the mouth up from 3 feet. The results can be seen in Table 8 where, for example, the first row "S&G 3/3/2 0/0/1 22/22/22" is interpreted to mean that the pair of preservice teachers S&G performed the experiments 25 times. The cup landed mouth down 3 times, 3 times, and 2 times for the three experiments. The cup landed mouth up 0 times, 0 times, and one time. Finally the cup landed on its side 22 times for each experiment.

The preservice teachers were struck by the remarkable consistency of the results and began to see that, although it might be hard to develop a theory, there is very strong evidence here that there is an observable underlying phenomenon at

In the discussion that followed this experiment, the preservice teachers were surprised at the stability and consistency of the results across groups.

Table 8 Data from sets of 25 cup drops with various initial positions

	Large O	Small O	Side
S & G	3/3/2	0/0/1	22/22/22
G & P	3/3/4	1/1/0	21/21/21
C & F	0/3/2	2/0/1	23/22/22
B & I	3/2/1	0/2/0	22/21/24
T & R	2/2/2	1/1/0	22/22/23
K & J	2/0/2	1/1/0	22/24/23
M & J	4/2/3	1/0/0	20/23/22
D & N	5/1/2	1/2/1	19/22/22
Total	22/16/18	7/7/3	171/177/179
Total%	11/8/9	3.5/3.5/1.5	88.5/88.5/89.5

play here. Furthermore they could see that experiments can predict those probabilities very closely. This is a powerful example of Monte Carlo reasoning (Rubinstein, 1981) whereby the probability is established by repeated experiment in the absence of theory and leads the preservice teachers to the important concept that even in the absence of a theory, the evidence of an experiment can result in the ability to make realistic predictions for other experiments.

Part V: We don't have a theory, and it's hard to imagine what one would look like

For the final activity, the preservice teachers were asked to clasp their hands and then record whether their left thumb or right thumb is on top. Everyone has a natural way of doing this as can be attested to if you try to reverse what you did the first time: it feels strange the “wrong” way. The results were: right on left 6 and left on right 15. The preservice teachers then discussed how this result might be explained: Is there a theory? Is there any way to know whether this class was representative or anomalous? One theory was that it may be related to “handedness” but the class had 19 right-

handed people and 2 left-handed, so it was hard to argue for a correlation. The preservice teachers argued that, perhaps, a very large population sample might provide evidence that there is an underlying stable result and that there may be some genetic explanation other than “handedness.” However, there was no consensus as to whether there was an underlying phenomenon at play.

Conclusion

The five activities described above are designed to scaffold the preservice teachers through a series of experiences that explore different aspects of the relationship between theoretical probability and experimental probability. Preservice teachers are given the opportunity to explore how experiments can be used to confirm theories or develop theories and they can also see the limitations of experiments depending on the plausibility of the existence of an underlying theory. The two key pedagogical moments in the set of activities occurred in Part III, where the initial theory did not fit with the evidence, and Part IV, where the results were robust across experimental conditions. In the former case, the importance of learners having articulated expectations so that they are invested in the outcome of the experiment and can experience “cognitive dissonance” (Festinger, 1957) was highlighted. It was clear that the motivation of the preservice teachers to calculate theoretical probabilities was heightened when the outcome of the game was different from what they expected. Many researchers (e.g., Zaslavsky, 2005) have written about the potential for learning that arises from such cognitive dissonance. The other key juncture was the idea developed, through the dropping of the cups, that relatively accurate and useful probabilistic data can be determined even when it is not possible

It was clear that the motivation of the preservice teachers to calculate theoretical probabilities was heightened when the outcome of the game was different from what they expected.

or practical to calculate a theoretical probability. The preservice teachers were genuinely surprised at the robustness of the results in the cup dropping activity and were able to see the value of experimental probability in examining real-life phenomena. Each of the activities focuses on a particular aspect of the relationship between experiment and theory but, taken as a whole, they provide a more coherent picture than would be provided by an isolated activity that simply reaches the conclusion “the larger your sample size the more accurate the result.” The discussions also allow preservice teachers to reflect on why theories are needed and how they might be developed.

This set of activities could be used with preservice teachers as above to improve their content knowledge in the area of probability as well as providing both a model of inquiry-based approaches and a forum for discussing pedagogical techniques involving hands-on activities. They could also be used in middle school classrooms to help students experience the power of probability experiments in examining real-life phenomena.

References

- Common Core State Standards Initiative. (2010). *Common Core State Standards for mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, 12, 217-230.
- Liu, Y., & Thompson, P. (2007). Teachers’ Understandings of Probability. *Cognition and Instruction*, 25(2-3), 113-160.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Rubinstein, R. (1981). *Simulation and the Monte Carlo Method*. New York: J. Wiley.
- Zaslavsky, O. (2005). Seizing the opportunity to create uncertainty in learning mathematics. *Educational Studies in Mathematics*, 60, 297-321.



**M I C H A E L
MEAGHER,**
mmeagher@
brooklyn.cuny.
edu, is an Assistant
Professor of
M a t h e m a t i c s
Education at Brooklyn College - CUNY.
His research interests include teacher
preparation for urban schools and the
use of advanced digital technologies
in the teaching and learning of
mathematics.